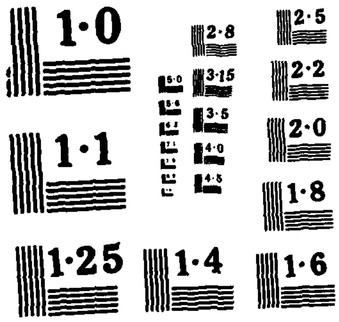
AD-A163 312 ON THE LEAST-INFORMATIVE DISTRIBUTION IN A CONTAMINATION NEIGHBORHOOD(U) JOHNS HOPKINS UNIV BALTIMORE MD DEPT OF MATHEMATICAL SCIENCES 1/1

UNCLASSIFIED J C WIERMAN ET AL. JUN 85 TR-434 F/G 12/1 NL

END



NATIONAL BUREAU OF STANDARDS
MICROCOPY RESOLUTION TEST CHART

On the Least-Informative Distribution in a Contamination Neighborhood

by

John C. Wierman

and

Kenneth S. Cantwell

The Johns Hopkins University
Baltimore, Maryland 21218
ONR Technical Report No. 85-2
Technical Report No. 434
May, 1985

AMS 1980 Subject Classification: 62F35

Key words and phrases: Robust estimation, Fisher information, contamination neighborhood.

Research supported by the U.S. Department of Navy under Office of Naval Research Contract No. N00014-79-C-0801 and the National Science Foundation under Grant No. MCS-8303238.

1. Introduction

The classical theory of robust estimation of a location parameter was developed by Huber (1964, 1981) in the following setting: Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution $F(x-\theta)$, where θ is an unknown location parameter. The distribution function F is not precisely known, but is assumed to be in a convex neighborhood $\mathcal{P}$ of a model distribution G. The goal is to estimate θ efficiently by a functional $T(\mathbb{F}_n)$, where $\mathbb{F}_n$ denotes the empirical distribution function associated with $X_1, X_2, \ldots, X_n$. For example, an M-estimate $T(\mathbb{F}_n)$ is a solution t of

$$0 = \int \psi(x-t)d\mathbb{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \psi(x_i-t)$$

for some estimating function ψ. Under suitable regularity conditions [see Huber (1967), Boos and Serfling (1980)], $T(\mathbb{F}_n)$ is a consistent estimator of θ, and $\sqrt{n}(T(\mathbb{F}_n)-\theta)$ is asymptotically normally distributed with mean zero and variance $V(\psi,F)$. One wishes to choose the estimating function ψ_0 which minimizes the maximum variance over all distributions in $\mathcal{P}$:

$$\sup_{F\in\mathcal{P}} V(\psi_0,F) = \inf \sup_{F\in\mathcal{P}} V(\psi,F).$$

Let F_0 be the distribution *minimizing* the Fisher information

$$I(F) = \int (f'/f)^2 f$$

over $\mathcal{P}$, where $f = F'$. Under general conditions there is a unique least-informative distribution in a convex neighborhood $\mathcal{P}$ of G [see Huber (1981), § 4.4] By the Cramér-Rao bound, the asymptotic variance of T at F_0 is at least $1/I(F_0)$. Finding a sequence of estimates $\{T_n\}$ for which the asymptotic variance is at most $1/I(F_0)$ for any $F\in\mathcal{P}$ then solves the minimax problem. Under suitable conditions, $\psi_0 = -f_0'/f_0$ is the desired minimax estimating

function, where $f_0 = F_0'$. Thus, a crucial step in this approach is the determination of the least informative distribution in an appropriate neighborhood of the model distribution. However, the least informative distribution is known only for a rather restricted class of model distributions and neighborhoods.

The theory is most completely developed for the ε-contamination neighborhood of the model distribution G, defined by

$$\mathcal{P}_\varepsilon(G) \equiv \{F = (1-\varepsilon)G + \varepsilon H\colon H \in \mathcal{M}\},$$

where $\mathcal{M}$ denotes the set of probability distributions on $\mathbb{R}$. Huber (1964) treated the location problem for symmetric distributions in the ε-contamination neighborhood of a model distribution G with a log-concave density g. The least-informative density is equal to $(1-\varepsilon)g$ on a central interval and is an exponential function on each of the remaining unbounded intervals. Huber's method of proof provides the motivation for our approach, and is sketched at the opening of section 2. Collins (1976) investigated a class of densities which were equal to the normal density on an interval of the form $[-d,d]$ and allowed asymmetry outside this interval.

The ε-Kolmogorov neighborhood of the standard normal distribution, defined by

$$\mathcal{P}_\varepsilon^K(\Phi) \equiv \{F \in \mathcal{M}\colon \sup_{x\in\mathbb{R}} |F(x) - \Phi(x)| \le \varepsilon\}$$

has been studied by Huber (1964), who determined the least favorable distribution for $\varepsilon < \varepsilon_0 \approx .0303$, and by Sachs and Ylvisaker (1972) for $\varepsilon \ge \varepsilon_0$.

This paper determines the form of the least informative distribution in the ε-contamination neighborhood for a broad class of model distributions. Since the assumption of a log-concave model density in Huber (1964) is

equivalent to strong unimodality of G, [Ibragimov (1956)], the assumption is quite restrictive. Our result is more broadly applicable, being valid for any model distribution G with a density which has a finite number of local maxima.

Let G denote a model distribution with density $g = G'$. For each $k > 0$, define

$$\mathcal{F}(k) \equiv \{h:\ h(x) = b\cosh^2[\tfrac{k}{2}(x-a)] \text{ and } h(x) \geq g(x)\ \forall\ x \in \mathbb{R}\}$$

and

$$C(k) \equiv \{x \in \mathbb{R}:\ \exists\ h \in \mathcal{F}(g,k) \text{ such that } h(x) = g(x)\},$$

and, for each $x \in \mathbb{R}$,

$$h_k(x) \equiv \inf\{h(x):\ h \in \mathcal{F}(g,k)\}.$$

Our principal result is the following:

<u>THEOREM</u>: Let G be a distribution with density g, such that $\psi_g \equiv -g'/g$ is continuously-differentiable and ψ_g' has finitely many intervals of increase. Then:

(i) For each $k > 0$, there exists an integer $N = N(k)$ and $-\infty \leq a_1 \leq b_1 < a_2 \leq b_2 < \ldots \leq b_N \leq +\infty$ such that

$$C(k) = (a_1, b_1] \cup \left\{ \bigcup_{i=2}^{N-1} [a_i, b_i] \right\} \cup [a_N, b_N).$$

(ii) For each $k > 0$, h_k has the form

$$h_k(x) = g(x) \qquad \text{if } x \in C(k)$$

$$= \left\{ \frac{\sqrt{g(b_i)}\sinh[\frac{k}{2}(a_{i+1}-x)] + \sqrt{g(a_{i+1})}\sinh[\frac{k}{2}(x-b_i)]}{\sinh[\frac{k}{2}(a_{i+1}-b_i)]} \right\}^2$$

$$\text{if } x \in [b_i, a_{i+1}],\quad 1 \leq i \leq N-1$$

$$= g(a_1)e^{k(x-a_1)} \qquad \text{if } x \le a_1 \text{ (if } a_1 \ne -\infty)$$

$$= g(b_N)e^{k(b_N-x)} \qquad \text{if } x \ge b_N \text{ (if } b_N \ne +\infty)$$

(iii) For each $\varepsilon > 0$, there exists $k = k(\varepsilon)$ such that $(1-\varepsilon)h_k$ is the probability density of the least informative distribution in $\mathcal{P}_\varepsilon(G)$ for the location problem.

The least informative density is reduced to $(1-\varepsilon)g$ at the peaks of g, so it is on the boundary of the neighborhood $\mathcal{P}_\varepsilon(G)$. The mass removed from the peaks is redistributed to fill in regions where the density g is steepest.

The pure scale parameter problem involves estimating σ for a family of densities $\frac{1}{\sigma} g(\frac{x}{\sigma})$, $\sigma > 0$. The Fisher information for scale is

$$I(F;\sigma) = \frac{1}{\sigma^2} \int \left[- \frac{g'(x)}{g(x)} x - 1 \right]^2 g(x)dx.$$

Since the scale problem can be transformed into the location problem by taking logarithms, the principal theorem also determines the form of the least informative distribution for the scale problem. [Alternatively, following Thall (1979), the results can be derived directly in the scale problem setting.] As in Huber (1981), §5.6, we consider symmetric distributions $F \in \mathcal{P}_\varepsilon(G)$, and obtain the following result:

<u>COROLLARY</u>: Let G be a distribution with density g, such that χ' has finitely many intervals of increase, where $\chi(x) = -x\psi_g(x)-1$. Then for each $\varepsilon > 0$, the least informative density in $\mathcal{P}_\varepsilon(G)$ for the scale problem has the form

$$f(x) = (1-\varepsilon)g(x) \qquad \text{if } t_i \le x \le u_i,\ i = 1,2,\ldots,N$$

$$= (1-\varepsilon)\left\{ \frac{\sqrt{t_{i+1}^k g(t_{i+1})}(x^k-u_i^k) + \sqrt{u_i^k\ g(u_i)}\ (t_{i+1}^k-x^k)}{x^{k/2}(t_{i+1}^k-u_i^k)} \right\}^2$$

$$\text{if } u_i \le x \le t_{i+1},\ i = 1,2,\ldots,N-1$$

$$= 2(1-\varepsilon)\left(\frac{|x|}{t_1}\right)^{k-1} g(t_1) \qquad \text{if } x \le t_1 \text{ (if } t_1 \ne -\infty)$$

$$= 2(1-\varepsilon)\left(\frac{u_N}{|x|}\right)^{k+1} g(u_N) \qquad \text{if } x \ge u_N \text{ (if } u_N \ne +\infty),$$

where $-\infty \le t_1 \le u_1 < t_2 \le u_2 < \ldots \le u_N \le +\infty$.

2. Proof

The Differential Equation and Solution: Since a key differential equation arises in Huber's condition for the least-informative density, we begin with a brief description of his method for log-concave g. Consider a distribution $F \in \mathcal{P}_\varepsilon(G)$ with density f, and suppose that $\psi = -f'/f$ is continuous and piecewise differentiable. Then F is the least informative distribution in $\mathcal{P}_\varepsilon(G)$ if and only if for some $k > 0$

$$(*) \qquad \int (k^2 + 2\psi' - \psi^2)(f_1 - f) \geq 0$$

for all $f_1 = F_1'$, where $F_1 \in \{F \in \mathcal{P}_\varepsilon(G) : I(F) < \infty\}$. If g is log-concave, this condition is satisfied by setting $f = (1-\varepsilon)g$ when $|\psi_g| < k$ to obtain both $f_1 \geq f$ and $k^2 + 2\psi' - \psi^2 > 2\psi_g' \geq 0$, and letting f be a multiple of e^{kx} or e^{-kx} on intervals where $|\psi_g| \geq k$ to obtain $k^2 + 2\psi' - \psi^2 = 0$.

An important observation is that, while log-concavity is sufficient to obtain $k^2+2\psi' - \psi^2 \geq 0$, it is not necessary. This suggests replacing $(1-\varepsilon)g$ by a density which satisfies the differential equation $k^2 + 2\psi' - \psi^2 = 0$ on an interval where $|\psi_g| < k$ and $k^2 + 2\psi_g' - \psi_g^2 < 0$.

The different equation system

$$k^2 + 2\psi' - \psi^2 = 0$$
$$\psi = -f'/f$$
$$|\psi| < k$$

may be solved by separation of variables, to obtain solutions of the form

$$\psi(z) = \psi(z;k,c) = -k \tanh[\tfrac{k}{2}(z-c)]$$
$$f(z) = f(z;k,c,d) = d \cosh^2[\tfrac{k}{2}(z-c)] = \frac{d}{2}\cosh[k(z-c)] + \frac{d}{2}.$$

where $d > 0$ and $c \in \mathbb{R}$. Note that for any fixed k, a function $f(\cdot;k,c,d)$ is uniquely determined by the values at two points.

The function $f(\cdot;k,c,d)$ passing through (a_1,b_1) and (a_2,b_2) where $a_1 < a_2$, is given by

$$\left[\frac{\sqrt{b_1}\,\sinh[\frac{k}{2}(a_2-z)] + \sqrt{b_2}\,\sinh[\frac{k}{2}(z-a_1)]}{\sinh[\frac{k}{2}(a_2-a_1)]}\right]^2 ,$$

the form employed in the statement of the Theorem.

For each $k > 0$ and $x \in \mathbb{R}$, let $g(\cdot;k,x)$ denote the function $f(\cdot;k,c,d)$ such that $f(x;k,c,d) = g(x)$ and $f'(x;k,c,d) = g'(x)$. We find that

$$g(\cdot;k,x) = f(\cdot;k,c(k,x),d(k,x))$$

where

$$c(k,x) = x - \frac{1}{k}\log\left(\frac{k-\psi_g(x)}{k+\psi_g(x)}\right)$$

and

$$d(k,x) = g(x)\left[1 - \frac{\psi_g(x)^2}{k^2}\right]$$

Note that c and d are continuous functions of k and x in the region where $|\psi_g(x)| < k$.

For each $x \in C(k)$, $g(\cdot;k,x)$ is the element of $F(g,k)$ that lies above g everywhere on $\mathbb{R}$ and is tangent to g at x.

For any $s < t$, and $s, t \in C(k)$, either $g(z;k,s) \leq g(z;k,t)\ \forall z \leq s$ and $g(z;k,t) \leq g(z;k,s)\ \forall z \geq t$, or $g(\cdot;k,t) = g(\cdot;k,s)$, since two points uniquely determine a function of this form.

<u>The Structure of C(k)</u>: Letting $\psi_{k,x}(\cdot) = -\,g'(\cdot;k,x)/g(\cdot;k,x)$, by tangency of g and $g(\cdot;k,x)$ at x we have

$$\psi_g(x) = \psi_{k,x}(x) = -\,k\tanh[\tfrac{k}{2}(x-c(k,x))]$$

so $|\psi_g(x)| < k$.

Note also that for $x \in C(k)$, $g''(x) \le g''(x;k,x)$ since $g(z) \le g(z;k,x)$ for all $z \in \mathbb{R}$. Therefore

$$\begin{aligned}\psi_g'(x) &= -g''(x)/g(x) + \psi_g^2(x)\\ &\ge -g''(x;k,x)/g(x;k,x) + \psi_{k,x}^2(x)\\ &= \psi_{k,x}'(x),\end{aligned}$$

so

$$k^2 + 2\psi_g'(x) - \psi_g^2(x) \ge k^2 + 2\psi_{k,x}'(x) - \psi_{k,x}^2(x) = 0.$$

Therefore

$$C(k) \subseteq B(k) \equiv \{x \in \mathbb{R}: |\psi_g(x)| < k \text{ and } k^2 + 2\psi_g'(x) - \psi_g^2(x) \ge 0\}.$$

Since ψ_g' has finitely many intervals of increase and decrease, for each $k > 0$, $B(k)$ is a finite union of intervals $B_i(k)$, $i = 1,2,\ldots,N(k)$.

For a fixed i, suppose that there exist $s, t \in C(k) \cap B_i(k)$, where $s < t$. Consider $x \in (s,t)$, and construct $g(\cdot;k,x)$. By construction, $\psi_{k,x}' = \frac{1}{2}(\psi_{k,x}^2 - k^2)$ on (s,t). On $(s,t) \subseteq B_i(k)$, $k^2 + 2\psi_g' - \psi_g^2 \ge 0$, which implies that $\psi_g' \ge \frac{1}{2}(\psi_g^2 - k^2)$. Since g and $g(\cdot;k,x)$ are tangent at x, $\psi_g(x) = \psi_{k,x}(x)$. Therefore $\psi_g(z) \ge \psi_{k,x}(z)$ for $x \le z \le t$, and $\psi_g(z) \le \psi_{k,x}(z)$ for $s \le z \le x$. Integrating, we obtain for $x \le y \le t$

$$\log \frac{g(x)}{g(y)} = \int_x^y \psi_g(z)\,dz \ge \int_x^y \psi_{k,x}(z)\,dz = \log \frac{g(x;k,x)}{g(y;k,x)},$$

so

$$g(y;k,x) \ge g(y) \quad \forall \quad x \le y \le t.$$

Similarly,

$$g(y;k,x) \ge g(y) \quad \forall \quad s \le y \le x.$$

For $y \leq s$,

$$g(y;k,x) \geq g(y;k,s) \geq g(y)$$

and for $y \geq t$,

$$g(y;k,x) \geq g(y;k,t) \geq g(y).$$

Therefore, $g(\cdot;k,x) \geq g$ on $\mathbb{R}$, so $x \in C(k)$.

Hence $C(k)$ is union of intervals $C_i(k)$, with $C_i(k) \subseteq B_i(k)$. [For some i, $C_i(k)$ may be empty.]

Suppose x is the right endpoint of $C_i(k)$, where $C_i(k) \neq \{x\}$. If $|\psi_g(x)| < k$, then by continuity of $c(k,x)$ and $d(k,x)$, we have for each $z \in \mathbb{R}$

$$g(z;k,x) = \lim_{y \uparrow x} g(z;k,y) \geq g(z),$$

so $x \in C_i(k)$.

If $|\psi_g(x)| = k$, then $x \notin C(k)$, since if g and $g(\cdot;k,x)$ were tangent,

$$|\psi_g| = |\psi_{k,x}| = |-k \tanh[\tfrac{k}{2}(x-c(k,x))]| < k.$$

Furthermore,

$$\lim_{y \uparrow x} c(k,y) = +\infty \text{ and } \lim_{y \uparrow x} d(k,y) = 0.$$

For $y_1, y_2 \in C_i(k)$, $y_1 < y_2 < x$, recalling that $g(z;k,y_1) \geq g(z;k,y_2)\ \forall\, z \geq y_2$, and expressing $g(\cdot;k,y)$ in terms of hyperbolic sine functions, we obtain monotone convergence at each $z > x$:

$$\lim_{y\uparrow x}\left[\frac{\sqrt{g(y)}\sinh[\frac{k}{2}(c(k,y)-z)] + \sqrt{d(k,y)}\sinh[\frac{k}{2}(z-y)]}{\sinh[\frac{k}{2}(c(k,y)-y)]}\right]^2 = g(x)e^{k(x-z)}$$

Thus, there is no $x_1 > x$ such that $x_1 \in C(k)$, since otherwise there exists $f(\cdot;k,c,d)$ which is either tangent to the exponential function or intersects it twice, in either case violating $|f'(\cdot;k,c,d)/f(\cdot;k,c,d)| < k$.

Therefore, $C(k)$ includes the right endpoints of all intervals $C_i(k)$ except possibly the rightmost. Similarly, $C(k)$ includes the left endpoints of all subintervals $C_i(k)$ except possibly the leftmost.

<u>The Form of h_k</u>: Note that

$$h_k(x) = \inf\{f(x): f \in F(g,k) \text{ and } f(y) = g(y) \text{ for some } y \in \mathbb{R}\}.$$

If $x \in C(k)$, clearly $h_k(x) = g(x)$. If $x \notin C(k)$, then $x \in (s,t) \subseteq C(k)^c$, where s and t are endpoints of intervals of $C(k)^c$.

We claim that if $[s,t]$ is bounded, then on $[s,t]$

$$h_k = g(\cdot;k,s) = g(\cdot;k,t).$$

In this case $s,t \in C(k)$ by the previous subsection. Thus, $g(t;k,s) \geq g(t)$ and $g(s;k,t) \geq g(s)$. Suppose the first of these inequalities is strict. Then there exists $\varepsilon > 0$ such that $\forall\, z \in \mathbb{R}$

$$g(z-\varepsilon;k,s) = d(k,s)\cosh^2[\tfrac{k}{2}(z-\varepsilon-c(k,s))] > g(z),$$

and thus for some $b < 1$, there exists $y \in (s,t)$ such that $bg(y-\varepsilon;k,s) = g(y)$ and $bg(z-\varepsilon;k,s) \geq g(z)\ \forall\, z \in \mathbb{R}$. This implies that $y \in C(k)$, which is a contradiction, so in fact $g(t;k,s) = g(t)$, and similarly $g(s;k,t) = g(s)$. Since two points determine a function of this form, $g(\cdot;k,s) = g(\cdot;k,t)$.

Suppose there exists $w \in (s,t)$ such that $h_k(w) < g(w;k,s) = g(w;k,t)$. Then for sufficiently small $\varepsilon > 0$, there exist c and d such that $f(\cdot;k,c,d) \geq g$ on $\mathbb{R}$ and

$$f(w;k,c,d) < h_k(w) + \varepsilon < g(w;k,s).$$

Since $f(s;,k,c,d) \geq g(s)$ and $f(t;k,c,d) \geq g(t)$, $f(\cdot;,k,c,d)$ intersects $g(\cdot;k,s)$ at two points, so the two functions are identical. This contradiction establishes that on $[s,t]$

$$h_k = g(\cdot;k,s) = g(\cdot;k,t).$$

If $[t,+\infty) \subseteq C(k)^c$, where t is an endpoint of $C_i(k)$ for some i, then $\psi_g(t) = k$. To see this, note that if $\psi_g(t) < k$, by continuity of $c(k,x)$ and $d(k,x)$ as above, $g(z;k,t) \geq g(z), \forall z \in \mathbb{R}$, and $g(\cdot;k,t)$ is not tangent to g at any point $z > t$. By translating $g(\cdot;k,t)$ to the right by a sufficiently small $\varepsilon > 0$, then reducing the coefficient until the resulting curve is tangent to g, we see that there exists $x > t$ with $x \in C(k)$, which is a contradiction.

Since $\psi_g(t) = k$, the reasoning in the previous subsection shows that, for $z > t$,

$$h_k(z) = \lim_{x \uparrow t} g(z;k,x) = g(t)e^{k(t-z)}$$

Similarly, if $(-\infty,s] \subseteq C(k)^c$, where s is an endpoint of $C_i(k)$ for some i, then for $z < s$, $h_k(z) = g(s)e^{k(z-s)}$.

Pointwise Convergence: A straightforward computation shows that for $k^* > k$, $x \in \mathbb{R}$, $g(z;k^*,x) \geq g(z;k,x)$ for all $z \in \mathbb{R}$. Therefore if $x \in C(k)$, then $x \in C(k^*)$ also.

Recall that $C(k) = \bigcup_{i=1}^{N(k)} C_i(k)$, where each $C_i(k)$ is an interval. We denote the endpoints of $C_i(k)$ by $a_i(k)$ and $b_i(k)$, with $a_i(k) \leq b_i(k)$ for all i and $b_i(k) < a_{i+1}(k)$ for $i = 1,\ldots,N(k)-1$. There are finitely many values

$$f(w;k,c,d) < h_k(w) + \varepsilon < g(w;k,s).$$

Since $f(s;,k,c,d) \geq g(s)$ and $f(t;k,c,d) \geq g(t)$, $f(\cdot;,k,c,d)$ intersects $g(\cdot;k,s)$ at two points, so the two functions are identical. This contradiction establishes that on $[s,t]$

$$h_k = g(\cdot;k,s) = g(\cdot;k,t).$$

If $[t,+\infty) \subseteq C(k)^c$, where t is an endpoint of $C_i(k)$ for some i, then $\psi_g(t) = k$. To see this, note that if $\psi_g(t) < k$, by continuity of $c(k,x)$ and $d(k,x)$ as above, $g(z;k,t) \geq g(z)$, $\forall z \in \mathbb{R}$, and $g(\cdot;k,t)$ is not tangent to g at any point $z > t$. By translating $g(\cdot;k,t)$ to the right by a sufficiently small $\varepsilon > 0$, then reducing the coefficient until the resulting curve is tangent to g, we see that there exists $x > t$ with $x \in C(k)$, which is a contradiction.

Since $\psi_g(t) = k$, the reasoning in the previous subsection shows that, for $z > t$,

$$h_k(z) = \lim_{x \uparrow t} g(z;k,x) = g(t)e^{k(t-z)}$$

Similarly, if $(-\infty,s] \subseteq C(k)^c$, where s is an endpoint of $C_i(k)$ for some i, then for $z < s$, $h_k(z) = g(s)e^{k(z-s)}$.

Pointwise Convergence: A straightforward computation shows that for $k^* > k$, $x \in \mathbb{R}$, $g(z;k^*,x) \geq g(z;k,x)$ for all $z \in \mathbb{R}$. Therefore if $x \in C(k)$, then $x \in C(k^*)$ also.

Recall that $C(k) = \bigcup_{i=1}^{N(k)} C_i(k)$, where each $C_i(k)$ is an interval. We denote the endpoints of $C_i(k)$ by $a_i(k)$ and $b_i(k)$, with $a_i(k) \leq b_i(k)$ for all i and $b_i(k) < a_{i+1}(k)$ for $i = 1,\ldots,N(k)-1$. There are finitely many values

of k at which $N(k)$ is discontinuous. Hence, for every $k^* > 0$, there exists $k^{**} > k^*$ such that $N(k)$ is constant for $k^* < k < k^{**}$ and $C_i(k) \subseteq C_i(k^{**})$ for all i for $k^* < k < k^{**}$.

By monotonicity of $C(k)$ in k, we have

$$\lim_{k \downarrow k^*} a_i(k) \leq a_i(k^*) \text{ and } \lim_{k \downarrow k^*} b_i(k) \geq b_i(k^*).$$

On the other hand, if $x \in C(k)$ for all $k > k^*$, by continuity of $g(\cdot;k,x)$ in k, then $x \in C(k^*)$. Thus,

$$\lim_{k \downarrow k^*} a_i(k) \geq a_i(k^*) \text{ and } \lim_{k \downarrow k^*} b_i(k) \leq b_i(k^*),$$

so equality holds. Since h_k depends continuously upon k and the endpoints of the intervals $C_i(k)$,

$$\lim_{k \downarrow k^*} h_k(z) = h_{k^*}(z) \quad \forall \; z \in \mathbb{R}.$$

We now consider pointwise convergence of h_k as k increases to k^*. As before, there exists $k^{**} < k^*$ such that $N(k)$ is constant for $k^{**} < k < k^*$.

Suppose $1 \leq i < N(k)$. Then for $b_i(k) < z < a_{i+1}(k)$,

$$h_k(z) = g(z;k,b_i(k)) = g(z;k,a_{i+1}(k)).$$

By monotonicity of $C(k)$, we have that

$$\lim_{k \uparrow k^*} b_i(k) = b_i \text{ and } \lim_{k \uparrow k^*} a_{i+1}(k) = a_{i+1}$$

exist, and therefore h_k converges pointwise monotonically to

$$H(z) \equiv g(z;k^*,b_i) = g(z;k^*,a_{i+1}) \; \forall \; b_i \leq z \leq a_{i+1}.$$

By continuity of $g(\cdot;k,x)$ in k and x, $g(\cdot;k,b_i) \geq g$ on $\mathbb{R}$, and is equal to g at b_i and a_{i+1}. Therefore

$$h_{k^*}(z) = H(z) = \lim_{k\uparrow k^*} h_k(z)$$

for $b_i \le z \le a_{i+1}$.

Since $\lim_{k\uparrow k^*} h_k = g = h_{k^*}$ on (a_i, b_i) for $1 \le i \le N(k)$, we have pointwise convergence of h_k to h_{k^*} on $(a_1, b_{N(k)})$. The argument is complete if $a_1 = -\infty$ and $b_{N(k)} = +\infty$.

Suppose $b_{N(k)} < +\infty$. Then for $z \ge b_{N(k)}(k)$,

$$h_k(z) = g(b_{N(k)}(k))e^{k(b_{N(k)}(k)-z)}$$

By monotonicity of $C(k)$,

$$\lim_{k\uparrow k^*} b_{N(k)}(k) = b_N \text{ exists}$$

and therefore

$$\lim_{k\uparrow k^*} h_k(z) = g(b_N)e^{k^*(b_N-z)}$$

By continuity of $g(x)e^{k(x-z)}$ in x and k, and the fact that $h_k \ge g$, $g(b_N)e^{k^*(b_N-z)} \ge g(z)$ on $[b_N, +\infty)$ with equality at b_N. Thus,

$$h_{k^*}(z) = g(b_N)e^{k^*(b_N-z)} = \lim_{k\uparrow k^*} h_k(z)$$

for all $z \ge b_N$. A similar argument applies if $a_1 > -\infty$, so h_k converges pointwise to h_{k^*} on $\mathbb{R}$.

<u>Continuity of the Integral</u>: Clearly h_k is integrable for each $k > 0$, since there are finitely many intervals in $C(k)$ and h_k is integrable on each interval of $C(k)$ and $C(k)^c$.

Using monotonicity of h_k in k to provide a dominating function, by pointwise convergence and the dominated convergence theorem, we obtain

$$\lim_{k\to k^*} \int_{\mathbb{R}} h_k = \int_{\mathbb{R}} h_{k^*}$$

for each $k^* > 0$. Thus, $\int_{\mathbb{R}} h_k$ is a continuous function of k.

We now show that $\lim_{k \downarrow 0} \int_{\mathbb{R}} h_k = +\infty$. Note that $|\psi_k| \le k$, or equivalently, $-k \le (-\log h_k)' \le k$. This implies that for any $x_0 \in \mathbb{R}$ and $t > 0$,

$$h_k(x_0 \pm t) \geq h_k(x_0)e^{-kt}.$$

Choose x_0 such that $g(x_0) = \max_{x \in \mathbb{R}} g(x)$, so $x_0 \in C(k)\ \forall k$. Then

$$\int_{\mathbb{R}} h_k(x)\,dx \geq 2\int_0^\infty h_k(x_0)e^{-kt}dt = \frac{2h_k(x_0)}{k} = \frac{2g(x_0)}{k} \longrightarrow \infty$$

as $k \to 0$.

For every $x \in \mathbb{R}$, $g(z;k,x) \geq g(z)\ \forall z \in \mathbb{R}$ for k sufficiently large, so $\lim_{k\to\infty} h_k(x) = g(x)$. Applying the dominated convergence theorem,

$$\lim_{k\to\infty} \int_{\mathbb{R}} h_k = \int_{\mathbb{R}} g = 1.$$

Therefore, for each $\varepsilon > 0$, there exists a $k(\varepsilon)$ such that

$$\int_{\mathbb{R}} h_{k(\varepsilon)}(t)\,dt = \frac{1}{1-\varepsilon}.$$

Then $(1-\varepsilon)h_{k(\varepsilon)}$ is the probability density function of a distribution in $P_\varepsilon(G)$.

<u>Least-Informative</u>: We now show that $f = (1-\varepsilon)h_{k(\varepsilon)}$ is the density of the least informative distribution in $P_\varepsilon(G)$, by checking condition (*). ψ is clearly piecewise differentiable, and at each endpoint x of $C(k(\varepsilon))$ tangency of $g(\cdot;k(\varepsilon),x)$ and g implies that $\psi_g(x) = \psi_{k(\varepsilon),x}(x)$, so ψ is continuous at x.

On $C(k(\varepsilon))$, $k(\varepsilon)^2 + 2\psi' - \psi^2 = k(\varepsilon)^2 + 2\psi_g' - \psi_g^2 \geq 0$ and $f_1 \geq f = (1-\varepsilon)g$, while on $C(k(\varepsilon))^c$, $k(\varepsilon)^2 + 2\psi' - \psi^2 = 0$, so

$$\int_{\mathbb{R}} (k(\varepsilon)^2 + 2\psi' - \psi^2)\ (f_1 - f) \geq 0$$

for any $f_1 = F_1'$ where $F_1 \in P_\varepsilon(G)$.

Acknowledgment: We express our gratitude to Robert J. Serfling for introducing us to a robust estimation problem which motivated the results of this paper, and for helpful discussions and suggestions. We acknowledge related preliminary work by Siu-on Lo, a former student of Dr. Serfling.

3. References

Boos, D.D., and Serfling, R.J. (1980) A note on differentials and the CLT and LIL for statistical functions, with application to M-estimates, Ann. Statist. 8, 618-624.

Collins, J.R. (1976) Robust estimation of a location parameter in the presence of asymmetry, Ann. Statist. 4, 68-85.

Huber, P.J. (1964) Robust estimation of a location parameter, Ann. Statist. 35, 73-101.

Huber, P.J. (1967) The behavior of maximum likelihood estimates under nonstandard conditions. Proc. Fifth Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1, University of California Press, Berkeley.

Huber, P.J. (1981) Robust Statistics, Wiley, New York.

Ibragimov, I.A. (1956) On the composition of unimodal distributions, Theor. Probab. Applic. 1, 255-260.

Sachs, J., and Ylvisaker, D. (1972) A note on Huber's robust estimation of a location parameter, Ann. Math. Statist. 43, 1068-1075.

Thall, P.F. (1979) Huber-sense robust M-estimation of a scale parameter, with application to the exponential distribution, JASA 74, 147-152.

Unclassfied

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT CATALOG NUMBER
ONR No. 85-2		

4. TITLE

On the Least-Informative Distribution in a Contamination Neighborhood

5. TYPE OF REPORT & PERIOD COVERED

Technical Report

6. PERFORMING ORGANIZATION REPORT NO.

Technical Report No. 434

7. AUTHOR(s)

John C. Wierman and Kenneth S. Cantwell

8. CONTRACT OR GRANT NUMBER(s)

ONR No. N00014-79-C-0801

9. PERFORMING ORGANIZATION NAME AND ADDRESS

Department of Mathematical Sciences
The Johns Hopkins University
Baltimore, Maryland 21218

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS

11. CONTROLLING OFFICE NAME & ADDRESS

Office of Naval Research
Statistics and Probability Research
Arlington, Virginia 22217

12. REPORT DATE

June, 1985

13. NUMBER OF PAGES

16

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)

15. SECURITY CLASS (of this report)

Unclassfied

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from report)

18. SUPPLEMENTARY NOTES

epsilon

19. KEY WORDS

Robust estimation, Fisher information, contamination neighborhood.

20. ABSTRACT

This document

For a broad class of model distributions G, we determines the distribution minimizing Fisher information for location in the ε-contamination neighborhood of G. The class consists of distributions G with a density g such that $\psi_g = -g'/g$ is continuously differentiable and ψ_g has finitely many intervals of increase. The result is extended to the scale problem.

psi sub g   psi sub g

Keywords, robust estimation